

Systems and Methods for Automated Analysis of Cells and Tissues

SUB A1

Under 35 USC §119(e)(1), this application claims the benefit of prior U.S. provisional application _____, filed April 19, 2001, and U.S. provisional application 60/334,723, filed October 31, 2001, the contents of which are incorporated herein by reference. Work described herein was supported in part by funding from the National Institute of Health. The U.S. Government may therefore have certain rights in the invention.

1. Background of the Invention

Tissue microarray technology offers the opportunity for high throughput analysis of tissue samples (Konen, J. et al., Nat. Med. 4:844-7 (1998); Kallioniemi, O.P. et al., Hum. Mol. Genet. 10:657-62 (2001); Rimm, D.L. et al., Cancer J. 7:24-31 (2001)). For example, the ability to rapidly perform large scale studies using tissue microarrays can provide critical information for identifying and validating drug targets/ prognostic markers (e.g. estrogen receptor (ER) and HER2/neu) and candidate therapeutics.

Automated quantitative analysis of tissue samples in microarrays, however, presents several challenges, including heterogeneity of tissue sections, subcellular localization of staining, and the presence of background signals. For example, depending on the type of tumor or tissue section being analyzed, the area of interest may represent nearly the entire sample, or only a small percentage. For instance, a pancreatic carcinoma or lobular carcinoma of the breast with substantial desmoplastic response may show stromal tissue representing a large percentage of the total area. If the goal of the assay is to determine epithelial cell expression of a given marker, a protocol must be used that evaluates only that region. The protocol must not only be able to select the region of interest but also normalize it, so that the expression level read from any given area can be compared with that of other areas. Subcellular localization presents similar challenges. Comparisons of nuclear or membranous staining, for example, are quite different from those in total cytoplasmic staining.

Certain methods (including confocal and convolution/ deconvolution microscopy) have been used to quantify expression of proteins at the cellular (or sub-cellular) level within a single high power field (Robinson, J.P. Methods Cell. Biol. 63:89-106 (2001); Shaw, P. Histochem. J. 26:687-94 (1994)). However, these are computationally intensive and laborious techniques, which operate on multiple serial images. As a result, the current standard for analysis of tissue microarrays, like tissue sections, is conventional pathologist-based analysis and grading of the sample according to scale.

Most biomarkers exhibit a parametric (normal, "bell-shaped") distribution, and consequently are best analyzed by a continuous scale (e.g., 0 to 1000). Unfortunately, manual observation tends to be nominal (e.g. 1+, 2+, 3+), primarily because the human eye is unable to reliably distinguish subtle differences in staining intensity. Several methods have been developed to translate nominal manual observations into a continuous scale. Foremost among these is the H-score where the percent of positively stained cells (0 to 100) is multiplied by the staining intensity (e.g. 0 to 3) to make a theoretically continuous scale (0 to 300). However, the inability to detect subtle differences in staining intensity, particularly at the low and high ends of the scale, as well as the tendency to round scores (e.g., 50% at 3+ for a score of 150, versus 47% at 3+ for a score of 141), limits the effectiveness of the H-score.

Automated systems and methods for rapidly analyzing tissue, including tissue microarrays, that permit the identification and localization of identified biomarkers within tissues and other cell containing samples, are needed.

2. Summary of the Invention

In one aspect, the invention features systems and methods for rapidly analyzing cell containing samples to localize and quantitate particular biomarkers within cells. In one embodiment, the method is implemented by a computer and superimposes an image of the biomarker against an image of a user defined area within the cell to determine whether the biomarker is within the user defined area.

In another aspect, the invention features an algorithm that facilitates the optical analysis of an array of biological samples, despite image irregularities, distortions, varying topologies, and the absence of one or more elements.

Analysis of patient samples according to the systems and processes described herein can be useful diagnostically (e.g. to identify patients who have a particular disease, have been exposed to a particular toxin or are responding well to a particular therapeutic or organ transplant) and prognostically (e.g. to identify patients who are likely to develop a particular disease, respond well to a particular therapeutic or be accepting of a particular organ transplant). As new and better markers of disease become identified in the post-genomic era, the instant described processes, which not only quantitate the markers, but also determine their relative location within a cell, will increase in applicability.

Automated analysis of cell containing preparations, as described herein, can provide a rapid assessment of the prognostic benefit of biomarkers. In addition, these automated techniques can identify associations that are typically not revealed using manual techniques. Also, automated analysis can better discern subtle differences in staining intensity, particularly at the upper and lower extremes. The ability to detect low level expression and distinguish it from no expression can provide important prognostic information. Furthermore, analysis of the sub-cellular distribution of certain biomarkers may elucidate previously unrecognized associations with patient survival.

Other features, objects, and advantages of the invention will be apparent from the following figures, detailed description and claims.

3. Description of the Figures

FIG. 1(A-D) shows separate monochromatic images of a colon carcinoma taken after staining with fluorescently-tagged markers and combined into a single color image as follows: DAPI (to visualize nuclei, blue), anti-cytokeratin (to distinguish tumor from non-tumor elements, green), and anti-alpha-catenin (to visualize cell membranes, red).

FIG. 2(A-D) shows a regression comparison of automated and pathologist-based scoring of estrogen receptor levels.

FIG. 3 is a flowchart of a method for identifying and accounting for the relative location of spots within an array.

FIG. 4 is a flowchart of a process for localizing a signal (e.g. a biomarker) within a locale.

5 FIG. 5 shows a tissue microarray.

FIG. 6 shows an optical microscope station.

4. Detailed Description

4.1 General

FIG. 3
FIG. 4
FIG. 5
FIG. 6
FIG. 7
FIG. 8
FIG. 9
FIG. 10
FIG. 11
FIG. 12
FIG. 13
FIG. 14
FIG. 15
FIG. 16
FIG. 17
FIG. 18
FIG. 19
FIG. 20
FIG. 21
FIG. 22
FIG. 23
FIG. 24
FIG. 25
FIG. 26
FIG. 27
FIG. 28
FIG. 29
FIG. 30
FIG. 31
FIG. 32
FIG. 33
FIG. 34
FIG. 35
FIG. 36
FIG. 37
FIG. 38
FIG. 39
FIG. 40
FIG. 41
FIG. 42
FIG. 43
FIG. 44
FIG. 45
FIG. 46
FIG. 47
FIG. 48
FIG. 49
FIG. 50
FIG. 51
FIG. 52
FIG. 53
FIG. 54
FIG. 55
FIG. 56
FIG. 57
FIG. 58
FIG. 59
FIG. 60
FIG. 61
FIG. 62
FIG. 63
FIG. 64
FIG. 65
FIG. 66
FIG. 67
FIG. 68
FIG. 69
FIG. 70
FIG. 71
FIG. 72
FIG. 73
FIG. 74
FIG. 75
FIG. 76
FIG. 77
FIG. 78
FIG. 79
FIG. 80
FIG. 81
FIG. 82
FIG. 83
FIG. 84
FIG. 85
FIG. 86
FIG. 87
FIG. 88
FIG. 89
FIG. 90
FIG. 91
FIG. 92
FIG. 93
FIG. 94
FIG. 95
FIG. 96
FIG. 97
FIG. 98
FIG. 99
FIG. 100
FIG. 101
FIG. 102
FIG. 103
FIG. 104
FIG. 105
FIG. 106
FIG. 107
FIG. 108
FIG. 109
FIG. 110
FIG. 111
FIG. 112
FIG. 113
FIG. 114
FIG. 115
FIG. 116
FIG. 117
FIG. 118
FIG. 119
FIG. 120
FIG. 121
FIG. 122
FIG. 123
FIG. 124
FIG. 125
FIG. 126
FIG. 127
FIG. 128
FIG. 129
FIG. 130
FIG. 131
FIG. 132
FIG. 133
FIG. 134
FIG. 135
FIG. 136
FIG. 137
FIG. 138
FIG. 139
FIG. 140
FIG. 141
FIG. 142
FIG. 143
FIG. 144
FIG. 145
FIG. 146
FIG. 147
FIG. 148
FIG. 149
FIG. 150
FIG. 151
FIG. 152
FIG. 153
FIG. 154
FIG. 155
FIG. 156
FIG. 157
FIG. 158
FIG. 159
FIG. 160
FIG. 161
FIG. 162
FIG. 163
FIG. 164
FIG. 165
FIG. 166
FIG. 167
FIG. 168
FIG. 169
FIG. 170
FIG. 171
FIG. 172
FIG. 173
FIG. 174
FIG. 175
FIG. 176
FIG. 177
FIG. 178
FIG. 179
FIG. 180
FIG. 181
FIG. 182
FIG. 183
FIG. 184
FIG. 185
FIG. 186
FIG. 187
FIG. 188
FIG. 189
FIG. 190
FIG. 191
FIG. 192
FIG. 193
FIG. 194
FIG. 195
FIG. 196
FIG. 197
FIG. 198
FIG. 199
FIG. 200
FIG. 201
FIG. 202
FIG. 203
FIG. 204
FIG. 205
FIG. 206
FIG. 207
FIG. 208
FIG. 209
FIG. 210
FIG. 211
FIG. 212
FIG. 213
FIG. 214
FIG. 215
FIG. 216
FIG. 217
FIG. 218
FIG. 219
FIG. 220
FIG. 221
FIG. 222
FIG. 223
FIG. 224
FIG. 225
FIG. 226
FIG. 227
FIG. 228
FIG. 229
FIG. 230
FIG. 231
FIG. 232
FIG. 233
FIG. 234
FIG. 235
FIG. 236
FIG. 237
FIG. 238
FIG. 239
FIG. 240
FIG. 241
FIG. 242
FIG. 243
FIG. 244
FIG. 245
FIG. 246
FIG. 247
FIG. 248
FIG. 249
FIG. 250
FIG. 251
FIG. 252
FIG. 253
FIG. 254
FIG. 255
FIG. 256
FIG. 257
FIG. 258
FIG. 259
FIG. 260
FIG. 261
FIG. 262
FIG. 263
FIG. 264
FIG. 265
FIG. 266
FIG. 267
FIG. 268
FIG. 269
FIG. 270
FIG. 271
FIG. 272
FIG. 273
FIG. 274
FIG. 275
FIG. 276
FIG. 277
FIG. 278
FIG. 279
FIG. 280
FIG. 281
FIG. 282
FIG. 283
FIG. 284
FIG. 285
FIG. 286
FIG. 287
FIG. 288
FIG. 289
FIG. 290
FIG. 291
FIG. 292
FIG. 293
FIG. 294
FIG. 295
FIG. 296
FIG. 297
FIG. 298
FIG. 299
FIG. 300
FIG. 301
FIG. 302
FIG. 303
FIG. 304
FIG. 305
FIG. 306
FIG. 307
FIG. 308
FIG. 309
FIG. 310
FIG. 311
FIG. 312
FIG. 313
FIG. 314
FIG. 315
FIG. 316
FIG. 317
FIG. 318
FIG. 319
FIG. 320
FIG. 321
FIG. 322
FIG. 323
FIG. 324
FIG. 325
FIG. 326
FIG. 327
FIG. 328
FIG. 329
FIG. 330
FIG. 331
FIG. 332
FIG. 333
FIG. 334
FIG. 335
FIG. 336
FIG. 337
FIG. 338
FIG. 339
FIG. 340
FIG. 341
FIG. 342
FIG. 343
FIG. 344
FIG. 345
FIG. 346
FIG. 347
FIG. 348
FIG. 349
FIG. 350
FIG. 351
FIG. 352
FIG. 353
FIG. 354
FIG. 355
FIG. 356
FIG. 357
FIG. 358
FIG. 359
FIG. 360
FIG. 361
FIG. 362
FIG. 363
FIG. 364
FIG. 365
FIG. 366
FIG. 367
FIG. 368
FIG. 369
FIG. 370
FIG. 371
FIG. 372
FIG. 373
FIG. 374
FIG. 375
FIG. 376
FIG. 377
FIG. 378
FIG. 379
FIG. 380
FIG. 381
FIG. 382
FIG. 383
FIG. 384
FIG. 385
FIG. 386
FIG. 387
FIG. 388
FIG. 389
FIG. 390
FIG. 391
FIG. 392
FIG. 393
FIG. 394
FIG. 395
FIG. 396
FIG. 397
FIG. 398
FIG. 399
FIG. 400
FIG. 401
FIG. 402
FIG. 403
FIG. 404
FIG. 405
FIG. 406
FIG. 407
FIG. 408
FIG. 409
FIG. 410
FIG. 411
FIG. 412
FIG. 413
FIG. 414
FIG. 415
FIG. 416
FIG. 417
FIG. 418
FIG. 419
FIG. 420
FIG. 421
FIG. 422
FIG. 423
FIG. 424
FIG. 425
FIG. 426
FIG. 427
FIG. 428
FIG. 429
FIG. 430
FIG. 431
FIG. 432
FIG. 433
FIG. 434
FIG. 435
FIG. 436
FIG. 437
FIG. 438
FIG. 439
FIG. 440
FIG. 441
FIG. 442
FIG. 443
FIG. 444
FIG. 445
FIG. 446
FIG. 447
FIG. 448
FIG. 449
FIG. 450
FIG. 451
FIG. 452
FIG. 453
FIG. 454
FIG. 455
FIG. 456
FIG. 457
FIG. 458
FIG. 459
FIG. 460
FIG. 461
FIG. 462
FIG. 463
FIG. 464
FIG. 465
FIG. 466
FIG. 467
FIG. 468
FIG. 469
FIG. 470
FIG. 471
FIG. 472
FIG. 473
FIG. 474
FIG. 475
FIG. 476
FIG. 477
FIG. 478
FIG. 479
FIG. 480
FIG. 481
FIG. 482
FIG. 483
FIG. 484
FIG. 485
FIG. 486
FIG. 487
FIG. 488
FIG. 489
FIG. 490
FIG. 491
FIG. 492
FIG. 493
FIG. 494
FIG. 495
FIG. 496
FIG. 497
FIG. 498
FIG. 499
FIG. 500
FIG. 501
FIG. 502
FIG. 503
FIG. 504
FIG. 505
FIG. 506
FIG. 507
FIG. 508
FIG. 509
FIG. 510
FIG. 511
FIG. 512
FIG. 513
FIG. 514
FIG. 515
FIG. 516
FIG. 517
FIG. 518
FIG. 519
FIG. 520
FIG. 521
FIG. 522
FIG. 523
FIG. 524
FIG. 525
FIG. 526
FIG. 527
FIG. 528
FIG. 529
FIG. 530
FIG. 531
FIG. 532
FIG. 533
FIG. 534
FIG. 535
FIG. 536
FIG. 537
FIG. 538
FIG. 539
FIG. 540
FIG. 541
FIG. 542
FIG. 543
FIG. 544
FIG. 545
FIG. 546
FIG. 547
FIG. 548
FIG. 549
FIG. 550
FIG. 551
FIG. 552
FIG. 553
FIG. 554
FIG. 555
FIG. 556
FIG. 557
FIG. 558
FIG. 559
FIG. 560
FIG. 561
FIG. 562
FIG. 563
FIG. 564
FIG. 565
FIG. 566
FIG. 567
FIG. 568
FIG. 569
FIG. 570
FIG. 571
FIG. 572
FIG. 573
FIG. 574
FIG. 575
FIG. 576
FIG. 577
FIG. 578
FIG. 579
FIG. 580
FIG. 581
FIG. 582
FIG. 583
FIG. 584
FIG. 585
FIG. 586
FIG. 587
FIG. 588
FIG. 589
FIG. 590
FIG. 591
FIG. 592
FIG. 593
FIG. 594
FIG. 595
FIG. 596
FIG. 597
FIG. 598
FIG. 599
FIG. 600
FIG. 601
FIG. 602
FIG. 603
FIG. 604
FIG. 605
FIG. 606
FIG. 607
FIG. 608
FIG. 609
FIG. 610
FIG. 611
FIG. 612
FIG. 613
FIG. 614
FIG. 615
FIG. 616
FIG. 617
FIG. 618
FIG. 619
FIG. 620
FIG. 621
FIG. 622
FIG. 623
FIG. 624
FIG. 625
FIG. 626
FIG. 627
FIG. 628
FIG. 629
FIG. 630
FIG. 631
FIG. 632
FIG. 633
FIG. 634
FIG. 635
FIG. 636
FIG. 637
FIG. 638
FIG. 639
FIG. 640
FIG. 641
FIG. 642
FIG. 643
FIG. 644
FIG. 645
FIG. 646
FIG. 647
FIG. 648
FIG. 649
FIG. 650
FIG. 651
FIG. 652
FIG. 653
FIG. 654
FIG. 655
FIG. 656
FIG. 657
FIG. 658
FIG. 659
FIG. 660
FIG. 661
FIG. 662
FIG. 663
FIG. 664
FIG. 665
FIG. 666
FIG. 667
FIG. 668
FIG. 669
FIG. 670
FIG. 671
FIG. 672
FIG. 673
FIG. 674
FIG. 675
FIG. 676
FIG. 677
FIG. 678
FIG. 679
FIG. 680
FIG. 681
FIG. 682
FIG. 683
FIG. 684
FIG. 685
FIG. 686
FIG. 687
FIG. 688
FIG. 689
FIG. 690
FIG. 691
FIG. 692
FIG. 693
FIG. 694
FIG. 695
FIG. 696
FIG. 697
FIG. 698
FIG. 699
FIG. 700
FIG. 701
FIG. 702
FIG. 703
FIG. 704
FIG. 705
FIG. 706
FIG. 707
FIG. 708
FIG. 709
FIG. 710
FIG. 711
FIG. 712
FIG. 713
FIG. 714
FIG. 715
FIG. 716
FIG. 717
FIG. 718
FIG. 719
FIG. 720
FIG. 721
FIG. 722
FIG. 723
FIG. 724
FIG. 725
FIG. 726
FIG. 727
FIG. 728
FIG. 729
FIG. 730
FIG. 731
FIG. 732
FIG. 733
FIG. 734
FIG. 735
FIG. 736
FIG. 737
FIG. 738
FIG. 739
FIG. 740
FIG. 741
FIG. 742
FIG. 743
FIG. 744
FIG. 745
FIG. 746
FIG. 747
FIG. 748
FIG. 749
FIG. 750
FIG. 751
FIG. 752
FIG. 753
FIG. 754
FIG. 755
FIG. 756
FIG. 757
FIG. 758
FIG. 759
FIG. 760
FIG. 761
FIG. 762
FIG. 763
FIG. 764
FIG. 765
FIG. 766
FIG. 767
FIG. 768
FIG. 769
FIG. 770
FIG. 771
FIG. 772
FIG. 773
FIG. 774
FIG. 775
FIG. 776
FIG. 777
FIG. 778
FIG. 779
FIG. 780
FIG. 781
FIG. 782
FIG. 783
FIG. 784
FIG. 785
FIG. 786
FIG. 787
FIG. 788
FIG. 789
FIG. 790
FIG. 791
FIG. 792
FIG. 793
FIG. 794
FIG. 795
FIG. 796
FIG. 797
FIG. 798
FIG. 799
FIG. 800
FIG. 801
FIG. 802
FIG. 803
FIG. 804
FIG. 805
FIG. 806
FIG. 807
FIG. 808
FIG. 809
FIG. 810
FIG. 811
FIG. 812
FIG. 813
FIG. 814
FIG. 815
FIG. 816
FIG. 817
FIG. 818
FIG. 819
FIG. 820
FIG. 821
FIG. 822
FIG. 823
FIG. 824
FIG. 825
FIG. 826
FIG. 827
FIG. 828
FIG. 829
FIG. 830
FIG. 831
FIG. 832
FIG. 833
FIG. 834
FIG. 835
FIG. 836
FIG. 837
FIG. 838
FIG. 839
FIG. 840
FIG. 841
FIG. 842
FIG. 843
FIG. 844
FIG. 845
FIG. 846
FIG. 847
FIG. 848
FIG. 849
FIG. 850
FIG. 851
FIG. 852
FIG. 853
FIG. 854
FIG. 855
FIG. 856
FIG. 857
FIG. 858
FIG. 859
FIG. 860
FIG. 861
FIG. 862
FIG. 863
FIG. 864
FIG. 865
FIG. 866
FIG. 867
FIG. 868
FIG. 869
FIG. 870
FIG. 871
FIG. 872
FIG. 873
FIG. 874
FIG. 875
FIG. 876
FIG. 877
FIG. 878
FIG. 879
FIG. 880
FIG. 881
FIG. 882
FIG. 883
FIG. 884
FIG. 885
FIG. 886
FIG. 887
FIG. 888
FIG. 889
FIG. 890
FIG. 891
FIG. 892
FIG. 893
FIG. 894
FIG. 895
FIG. 896
FIG. 897
FIG. 898
FIG. 899
FIG. 900
FIG. 901
FIG. 902
FIG. 903
FIG. 904
FIG. 905
FIG. 906
FIG. 907
FIG. 908
FIG. 909
FIG. 910
FIG. 911
FIG. 912
FIG. 913
FIG. 914
FIG. 915
FIG. 916
FIG. 917
FIG. 918
FIG. 919
FIG. 920
FIG. 921
FIG. 922
FIG. 923
FIG. 924
FIG. 925
FIG. 926
FIG. 927
FIG. 928
FIG. 929
FIG. 930
FIG. 931
FIG. 932
FIG. 933
FIG. 934
FIG. 935
FIG. 936
FIG. 937
FIG. 938
FIG. 939
FIG. 940
FIG. 941
FIG. 942
FIG. 943
FIG. 944
FIG. 945
FIG. 946
FIG. 947
FIG. 948
FIG. 949
FIG. 950
FIG. 951
FIG. 952
FIG. 953
FIG. 954
FIG. 955
FIG. 956
FIG. 957
FIG. 958
FIG. 959
FIG. 960
FIG. 961
FIG. 962
FIG. 963
FIG. 964
FIG. 965
FIG. 966
FIG. 967
FIG. 968
FIG. 969
FIG. 970
FIG. 971
FIG. 972
FIG. 973
FIG. 974
FIG. 975
FIG. 976
FIG. 977
FIG. 978
FIG. 979
FIG. 980
FIG. 981
FIG. 982
FIG. 983
FIG. 984
FIG. 985
FIG. 986
FIG. 987
FIG. 988
FIG. 989
FIG. 990
FIG. 991
FIG. 992
FIG. 993
FIG. 994
FIG. 995
FIG. 996
FIG. 997
FIG. 998
FIG. 999
FIG. 1000

In general, described herein are a collection of techniques that can be used for rapid, automated analysis of cell containing samples, including tissues and tissue microarrays. While these techniques build on one another and are described as a cohesive process, each technique has wide applicability and may be used individually or in combinations other than those described below.

In one embodiment is featured a technique to identify the location of spots within an image. The technique, termed “spotfinder”, can flexibly identify such locations despite image irregularities, distortions, varying topologies, and the absence of one or more elements. Although the process is described for locating the position of histospots and identifying missing histospots within tissue microarray images, the technique has broader application. More specifically it can be used to locate elements and identify missing elements in any collection of elements. Moreover, the process can be used on arrays of virtually any dimension and comprising a variety of elements. The specimens are not limited by size or shape, nor must they be regularly spaced.

In another embodiment is featured a technique that can be used alone or in conjunction with spotfinder to optically localize and quantitate a biomarker within a cell. Though an image of a cellular preparation typically features two dimensions, cellular preparations feature depth. For example, one cellular feature may rest atop another. This overlap can potentially confuse image analysis software. A technique described herein, dubbed RESA (Rapid Exponential

Subtraction Algorithm), can approximate a three dimensional image by subtracting out-of-focus image elements. Thus, the impact of background features on an image can be reduced, permitting better image analysis.

Another technique described herein, dubbed PLACE (Pixel Based Locale Assignment
5 Compartmentalization of Expression), distinguishes between different cellular characteristics. For example, the technique can determine the location of subcellular compartments within individual cells. A computer implementing this technique can, for instance, measure the relative intensities of images derived from compartment-specific stains on a pixel-by-pixel basis. The computer then determines for individual pixels within an image, the likelihood that the pixel
10 corresponds to a particular locale or user defined area within the cell. Such analysis permits the computer to assign signals to a sub-cellular compartment with an established degree of accuracy (e.g., 95%). The technique can co-localize signals associated with particular biomarkers with images of defined locales within cells.

Use of these techniques can enhance both the speed and accuracy of automated
15 microarray analysis. Figure 1 shows separate monochromatic images of a colon carcinoma taken after staining with fluorescently-tagged markers and combined into a single color image as follows: DAPI (to visualize nuclei, blue), anti-cytokeratin (to distinguish tumor from non-tumor elements, green), and anti-alpha-catenin (to visualize cell membranes, red) (panel A). Note the significant degree of overlap between the subcellular compartments. A monochromatic image of
20 the biomarker β -catenin, is taken (panel B, inset) and the intensity of each pixel in the image redistributed according to the relative signal intensity of the various compartments in panel A (blue= nuclei, red= membrane, green = cytoplasm).

Although the β -catenin expression in this tumor is predominantly membrane-associated, the significant overlap in compartments in panel A incorrectly assigns a significant amount of the
25 signal to the nucleus (magenta and blue pixels, panel B). To aid in the removal of overlapping signals, the monochromatic image of each sub-cellular compartment is exponentially subtracted from an out-of-focus image. Panel C shows a composite of the exponentially subtracted images of DAPI and anti-alpha-catenin (blue and red, respectively), shown on a mask derived from the anti-cytokeratin mask (green pixels). Pixels with too much overlap between channels are
30 negated (<5%), as are non-tumor areas, as defined by a mask generated from the anti-cytokeratin

image. In panel D, the signal intensity from an exponentially subtracted image of the biomarker (β -catenin, inset) is then redistributed according to the compartments defined in panel C. This results in more accurate assignment of the biomarker to the membrane compartment, which can have important prognostic significance. Since membrane-associated beta-catenin stabilizes cadherin-mediated adhesion by facilitating the cytoskeletal attachment of adhesion complexes, while nuclear-associated beta-catenin acts as a transcription factor, which up-regulates several genes important in cell proliferation and invasion and is considered an oncogene in this capacity, expression of beta-catenin alone does not provide prognostic information. However its localization in the nucleus can be an important indicator of carcinogenesis.

1. *Spot-finder*

In one embodiment, shown in FIG. 3, the computer removes B1 any atypically sized spots from the image. Atypically sized spots may include, for example, images of fused spots and/or debris. The computer performs the process automatically, though in other embodiments it may allow use of user input to facilitate the process.

The computer then creates or accesses an opaque virtual mask that is the size and shape of a typical spot. Using the virtual mask, the computer scans B2 the image to determine B3 where the mask first covers an area with the highest average pixel intensity. The computer monitors the total intensity of the image during the scan and, because the mask is opaque, identifies the position of the mask when the total image intensity is minimized. The computer identifies this area as the first spot and sets B4 the pixels within this area to have zero intensity. The computer also sets additional pixels within a predefined area around this area to have zero intensity. This helps to differentiate between overlapping spots.

After identifying the first spot, the computer again scans B2 the image using the mask to find the next area with the highest average pixel intensity. When the next area is found, the computer identifies it as the second spot and sets the pixels in and surrounding this area to have zero intensity. The computer repeats this process until it can no longer find areas of the image with sufficient intensity to qualify as spots.

The computer then identifies B5 a reference point (e.g., the center) in each spot, and draws a line connecting the reference point of each spot to each nearest neighboring spot reference point, above, below, to the left, and to the right. If the computer cannot identify a nearest neighbor in any of these directions (i.e., the spot is on the edge of the array), the computer draws a line from the center of the spot to the nearest edge of the image.

2. RESA and PLACE

Once the location of an image area of interest is determined, an optical microscope can obtain a high resolution image at an appropriate wavelength to identify cellular features of interest. These features include the biomarker, also referred to as the “signal”, the cells of interest within the tissue section (referred to as the “cell mask”), or a user defined location within the cell mask, also referred to as the “locale”. The signal, the cell mask, and the locale are referred to as “channels”.

Referring to FIG. 4, a process 50 determines the region of interest in the images by developing a mask from the cell mask channel (step C1). Next, the process applies this mask to the locale and signal channels (step C2). The process then removes out-of-focus information from the masked images, for example, in a process of pseudo-deconvolution (step C3). Next, in a “pixel assignment” phase, the process identifies subcellular features in the image, assigning pixels in the image to the locales (step C4). Once the pixels are assigned, the computer maps the locales onto the signal image (step C5), and quantifies the amount of biomarker in each locale. This phase is referred to as “signal assignment”. These steps are described in greater detail below.

Masking

During this process, the software identifies a region of interest in the image of the stained cells of interest (i.e., the cell mask channel). The software masks the locale and signal channels avoiding unnecessary analyses of areas outside the region of interest.

To identify a region of interest, the computer determines a threshold intensity for the cell mask channel. Once determined, the computer redistributes the pixel intensities in a binary redistribution. In other words, the computer sets the intensity of each pixel below the threshold

Pixels with intensity at or above the threshold intensity are assigned to the mask.

The mask is then further modified according to user-defined parameters and image processing techniques. For example, the mask can be dilated or eroded so that the mask area matches a user-defined value, or have holes of a particular (user-defined) size within it filled.

5 The user-defined parameters for creating the mask may be adjusted after analyzing a small number of sample histospot images, prior to running the entire array.

After developing the mask, the computer applies the mask to the images in the image stack, identifying the region of interest in each of these images as the pixel locations corresponding to the mask pixel locations.

10 Background Reduction

As shown, the process 50 reduces C3 the impact of the out-of-focus information from the image. For example, the process 50 may use a pseudo deconvolution technique. While the pixels of the remaining image are reduced in intensity, the image information represents a thinner virtual slice through the top of the tissue. Furthermore, pseudo-deconvolution enhances the interfacial areas between the higher stain intensity and lower stain intensity areas of the image by increasing the contrast between these areas.

15 The computer performs pseudo-deconvolution on the locale (cellular compartments) and signal (i.e., cellular components) channels. The computer first masks the images of these channels, reducing the number of pixels to be analyzed. The computer analyzes two images of each channel. The first image is an in-focus image (i.e., an image of the top of the histospot). The second image is a slightly out-of-focus image, produced by placing the focal plane slightly below the bottom of the tissue (e.g., for a five micron thick histospot, the focal plane of this image is located about eight microns below the top of the histospot).

25 For each pixel location, the computer subtracts a percentage of the out-of-focus image pixel intensity, $I_{out-of-focus}$, from the corresponding in-focus image pixel intensity, $I_{in-focus}$. The computer determines the adjusted pixel intensity, $I_{new\ pixel}$, using the quartile distribution, $D(\frac{1}{4})$, of the in-focus image as follows:

$$I_{\text{new pixel}} = I_{\text{in-focus}} - I_{\text{out-of-focus}} \times \left(\frac{I_{\text{max}} - I_{\text{in-focus}}}{I_{\text{max}}} \right)^{\psi},$$

where I_{max} is the maximum pixel intensity (e.g., 255 for an 8-bit image), and ψ is calculated from

$$\psi = \alpha \times D(Y_4)^{-\beta},$$

which was developed from an empiric assessment of a library of images. Optical deconvolutions were judged visually and the ψ for each was plotted versus the quartile distribution for the in-focus image. Regression analysis of the empiric data yielded values for the fitting-parameters (i.e., α is about 80 and β is about 1.19). The quartile distribution is determined from

$$D(Y_4) = \langle I \rangle_{\text{top}25} - \langle I \rangle_{\text{bottom}25},$$

where $\langle I \rangle_{\text{top}25}$ is the mean pixel intensity of the pixels within the top 25th percentile, and $\langle I \rangle_{\text{bottom}25}$ is the mean pixel intensity of pixels in the bottom 25th percentile. Conceptually, low intensity pixels in images with a low $D(Y_4)$ (i.e. a low signal to noise ratio) are subtracted less heavily than low intensity pixels from images with a high $D(Y_4)$.

The value of ψ may be refined by determining the percent of signal intensity remaining after pseudo-deconvolution within the masked area and comparing it to a predefined value for that channel. If the percent is, for example, greater than the predefined value then the pseudo-deconvolution stops. Otherwise, the computer iteratively increases the value of ψ until the predefined percent of signal intensity is reached. The predefined value is the expected percentage of the mask covered by a channel.

After pseudo-deconvolution, each pixel of the resulting images is assigned to a locale in a process referred to as pixel assignment.

Pixel Based Locale Assignment Compartmentalization of Expression (PLACE)

During the pixel assignment phase, the computer assigns an identity based on the relative intensity of that pixel location in each of the locale channel images (i.e., the images of the stained

locales). For example, during this phase the computer decides for each pixel location in the image whether it belongs to the nucleus, the membrane, or the cytoplasm. The computer does not make an assignment to pixels that it cannot assign within a user-defined degree of confidence (e.g., 95%). Higher levels of confidence eliminate more pixels from the analysis.

5 In general, for each pixel location in two locale images the computer reads a pixel intensity and compares each intensity value to a predetermined threshold intensity value. If the intensity value of only one locale is greater than the threshold, the computer assigns the pixel location to that locale. If both the intensity values are greater than their respective thresholds, the computer compares the intensity values from each locale, and assigns the identity of the locale having the greater intensity to that pixel location. If both the pixel intensities are below their threshold values, the computer assigns the pixel to a third locale.

10 After repeating the above for pixel locations in the images, the computer calculates the area of each locale, and compares the result to a predetermined (expected) coverage fraction. If the calculated coverage fraction (e.g., number of nuclear locale pixels/number of masked pixels) is greater than the predetermined coverage fraction, then the computer removes the pixels having the lowest intensity from the locale. The computer continues to remove the lowest intensity pixels until the coverage fraction is reduced to about the predetermined coverage fraction.

15 The following is an example of how this process works. The membrane locale and the nucleus locale images are selected for assignment analysis performed at 95% confidence interval. Pixel locations are assigned to the cytoplasm locale by exclusion.

20 The computer reads pixel intensities at each pixel location in the membrane and nucleus locale images, and compares them to threshold values. If neither of the intensity values are greater than the threshold values, the pixel location is assigned to the cytoplasm locale. If only the nuclear intensity or membrane intensity is greater than the threshold value, the computer assigns the pixel location to the above-threshold locale. If both intensities are higher than the thresholds, computer compares the ratio of the intensity values to one, and makes an assignment as follows:

$$\frac{\text{nuclear intensity}}{\text{membrane intensity}} > 1 \rightarrow \text{pixel location} = \text{nuclear locale}$$

$$\frac{\text{nuclear intensity}}{\text{membrane intensity}} < 1 \rightarrow \text{pixel location} = \text{membrane locale}.$$

$$\frac{\text{nuclear intensity}}{\text{membrane intensity}} = 1 \rightarrow \text{pixel location} = \text{unassigned}$$

Thus, if the nuclear intensity is greater than the membrane intensity, the computer assigns the pixel location to the nuclear locale. If the membrane intensity is greater than the nuclear intensity, the computer assigns the pixel location to the membrane locale. If the membrane intensity is equal to the nuclear intensity, the pixel location is unassigned. This repeats for the pixel locations.

Once all the pixel locations have been analyzed, the computer determines the amount of nuclear intensity incorrectly assigned to the membrane locale (i.e., nuclear to membrane spillover), and vice versa. If the amount of nuclear intensity incorrectly assigned to the membrane channel is >5% of the total nuclear intensity, then the computer weights the nuclear intensity by a factor, w , and recalculates the ratio of weighted nuclear intensity to membrane intensity. This ratio is compared to one, and pixel locations are reassigned as follows:

$$\frac{w \times \text{nuclear intensity}}{\text{membrane intensity}} > 1 \rightarrow \text{pixel location} = \text{nuclear locale}$$

$$\frac{w \times \text{nuclear intensity}}{\text{membrane intensity}} < 1 \rightarrow \text{pixel location} = \text{membrane locale}.$$

$$\frac{w \times \text{nuclear intensity}}{\text{membrane intensity}} = 1 \rightarrow \text{pixel location} = \text{unassigned}$$

The computer again determines the amount of each locale incorrectly assigned. If this is still >5% the computer increases the value of w and reiterates the steps above. This continues until the amount of incorrectly assigned nuclear locale is <5%. The computer employs a similar technique to minimize the membrane-to-nuclear spillover.

The computer also calculates the area of the cytoplasmic (exclusion) locale and compares it to a predetermined value. By iterating the assignment process, the computer ensures that there is <5% cytoplasmic-to-nuclear or cytoplasmic-to-membrane, as determined based on the biology.

The computer then evaluates the amount of signal in each locale during a “signal assignment” process.

Following pixel assignment, the computer sums the signal in each locale. The computer reads the pixel intensity of the signal image (i.e., the image of the stain that selectively labels the cellular component), and adds together the signal intensity for pixel locations assigned to like subcellular compartments. The computer calculates a pixel intensity sum of a locale by the direct addition of the signal intensity of each pixel location assigned to that locale. The computer also calculates a sum of pixel intensity ratios by adding together the ratio of the signal intensity and the locale intensity for each pixel location.

The pixel intensity sum and pixel intensity ratio sum is then used in calculating one or more parameters. For example, the computer determines the relative percentage of signal falling within each of the compartments (e.g. 30% of the total signal is membranous, 20% is cytoplasmic, and 50% is nuclear). In another example, the computer expresses the amount of signal present relative to the size of a particular compartment (e.g. the signal intensity of pixels assigned to the membrane channel divided by the number of pixels assigned to the membrane channel). The user may select to have the computer evaluate other parameters of interest. For example, how much of the image area is covered by the mask, how much of the mask is covered by each locale, etc.

By implementing the pseudo-deconvolution algorithm (which limits the majority of extraneous pixel intensity) together with intensity area measurements (which further define the area of a particular sub-cellular locale), the computer is able to make highly accurate assignments of pixel locations sub-cellular locations.

In some embodiments, the computer performs additional steps to better utilize the dynamic range of the camera. This is achieved by redistributing the pixel intensities in an image across the dynamic range of the detector based on their relative intensities.

One form of redistribution is normalized redistribution, whereby the lower threshold (i.e., the pixel intensity of the background, determined during masking) is subtracted from all the pixels in the image, and any pixel with a resulting negative value is set to zero. Normalized

redistribution is used for the signal channel as this redistribution preserves the scale from one sample to the next, allowing direct comparisons to be made between samples. This is performed either after masking the signal image.

Double-logarithmic redistribution sets all pixels in an image above 50% of the image's upper threshold (i.e., the value which only 50% of the pixels in the image have greater intensity) to the maximum intensity value (e.g., 255 for an 8-bit image). All pixels with intensity values below the lower threshold are set to 0, and all pixels with intensity values between the upper and lower thresholds are reassigned according to the formula:

$$I_{\text{new}} = \frac{\log(I_{\text{old}} - LT)}{I_{\text{max}} \log(\frac{1}{2}UT - LT)},$$

where I_{new} refers to the new pixel intensity, I_{old} refers to the old pixel intensity, LT and UT are the lower and 50% maximum thresholds, respectively, and I_{max} is the maximum intensity value.

Double-logarithmic redistribution is used for the locale channels, either after masking or after pseudo-deconvolution of these channels. Conceptually, it ensures that pixels in locale images that have intensities above the 50th percentile are assigned to their locale during the assignment phase. Pixels with intensities below, but close to, the 50th percentile are weighted more heavily and are more likely to be assigned to the locale than pixels that have intensities well below the 50th percentile.

Other user-defined redistributions, such as linear redistributions or other equation-based redistributions, may be used in addition to the above-described examples.

Although the algorithms described above are with reference to analysis of tissue microarrays, they are not limited to studying only such arrays. The spotfinder algorithm may be used for identifying the location of any element comprising a collection and the RESA and PLACE algorithms may be used to localize and quantitate a biomarker within any imageable, cell containing sample, including tissue biopsies and cell containing fluid samples, such as blood, urine, spinal fluid, saliva, lymph, pleural fluid, peritoneal fluid and pericardial fluid.

Also, any optical or non-optical imaging device can be used, such as for example, upright or inverted optical microscopes, scanning confocal microscopes, cameras, scanning or tunneling electron microscopes, scanning probe microscopes, and imaging infrared detectors etc.

5 In the embodiments described above, the computer can include hardware, software, or a combination of both to control the other components of the system and to analyze the images to extract the desired information about the histospots and tissue microarrays. The analysis described above is implemented in computer programs using standard programming techniques. Such programs are designed to execute on programmable computers each comprising a processor, a data storage system (including memory and/or storage elements), at least one input device, at least one output device, such as a display or printer. The program code is applied to input data (e.g., stitched together images or image stacks) to perform the functions described herein and generate information (e.g., localization of signal), which is applied to one or more output devices. Each computer program can be implemented in a high-level procedural or object-oriented programming language, or an assembly or machine language. Each such computer program can be stored on a computer readable storage medium (e.g., CD ROM or magnetic diskette) that when read by a computer can cause the processor in the computer to perform the analysis described herein.

10 The following provides a detailed description of a specific embodiment of the preparation and analysis of tissue microarrays according to methods described herein, although similar steps could be performed with respect to any cell containing sample. Referring to FIG. 5, a tissue microarray 100 includes multiple samples of histospots 120 prepared from histocores embedded typically in a thin (e.g., about five microns) block of paraffin 130 at regular intervals, forming a series of rows and columns. Histospots (thin sections of histocores) 120 may be substantially disk-like in shape and will typically have the same thickness as the paraffin block 130 (i.e., about 15 five microns) and a diameter of about 0.6 millimeters. Typically the centers of the histospots are spaced about a few tenths of a millimeter apart. Paraffin block 130 and histospots 120 may be mounted on a microscope slide 110. A tissue microarray 100 may include any number of histospots, typically on the order of several hundred to a few thousand.

20 Referring to FIG. 6, an optical microscopy station can be used to obtain an appropriate image of the tissue. Microscopy station 200 includes an inverted optical microscope 201 for 30

imaging the tissue, and a computer 290 for analyzing the images. Optical microscope 201 includes a mount 210, housing a light source 220, a sample stage 240, an objective lens 250 and a CCD camera 270. A frame grabber in computer 290 acquires the images through CCD camera 270.

5 Optical microscope 201 also includes filter wheels 230 and 260, which house a series of dichroic filters. The filters in wheel 230 allow selection of the appropriate illumination spectra for standard or fluorescent microscopy. Filters in wheel 260 filter the transmitted light for isolation of spectral signatures in fluorescent microscopy. Sample stage 240 supports and appropriately positions tissue microarray 100. Sample stage 240 can be linearly translated in the x, y, and z directions (axes are shown). Sample stage 240 includes motors to enable automated translation. Computer 290 controls sample stage 240 translation by servo control of the motors.

10 A tissue microarray 100 can be imaged as follows: a user places the microarray on a sample stage 240. The user adjusts sample stage 240 so that the first (i.e., top-left) histospot is at the center of the field of view and focused on CCD camera 270. The objective lens 250 should be adjusted to the appropriate resolution, for example, a 0.6 millimeter histospot can be viewed at 10X magnification. Generally, the histospots correspond to areas of higher light intensity than the surrounding paraffin, as assessed through various means including signals derived from the visible light scattering of stained tissues, tissue autofluorescence or from a fluorescent tag.

15 Computer 290 can acquire a low-resolution image (e.g. 64 pixel x 64 pixel with 16 bin resolution) using computer software (Softworx 2.5, Applied Precision, Issaquah, WA) and an imaging platform (e.g., Deltavision). Computer 290 automatically translates sample stage 240 by an amount approximately equal to a field of view. The computer then acquires a second low-resolution image. This process is repeated until the computer has acquired images of the entire tissue microarray. Then, using commercially available software, the computer generates a
20 composite image of the entire tissue microarray by stitching together the sequence of images like a patchwork.
25

30 Biological markers, which may be detected in accordance with the present invention include, but are not limited to any nucleic acids, proteins, peptides, lipids, carbohydrates or other components of a cell. Certain markers are characteristic of particular cells, while other markers have been identified as being associated with a particular disease or condition. Examples of

known prognostic markers include enzymatic markers such as galactosyl transferase II, neuron specific enolase, proton ATPase-2, and acid phosphatase. Hormone or hormone receptor markers include human chorionic gonadotropin (HCG), adrenocorticotrophic hormone, carcinoembryonic antigen (CEA), prostate-specific antigen (PSA), estrogen receptor, progesterone receptor, androgen receptor, gC1q-R/p33 complement receptor, IL-2 receptor, p75 neurotrophin receptor, PTH receptor, thyroid hormone receptor, and insulin receptor.

Lymphoid markers include alpha-1-antichymotrypsin, alpha-1-antitrypsin, B cell marker, bcl-2, bcl-6, B lymphocyte antigen 36kD, BM1 (myeloid marker), BM2 (myeloid marker), galectin-3, granzyme B, HLA class I Antigen, HLA class II (DP) antigen, HLA class II (DQ) antigen, HLA class II (DR) antigen, human neutrophil defensins, immunoglobulin A, immunoglobulin D, immunoglobulin G, immunoglobulin M, kappa light chain, kappa light chain, lambda light chain, lymphocyte/histocyte antigen, macrophage marker, muramidase (lysozyme), p80 anaplastic lymphoma kinase, plasma cell marker, secretory leukocyte protease inhibitor, T cell antigen receptor (JOVI 1), T cell antigen receptor (JOVI 3), terminal deoxynucleotidyl transferase, unclustered B cell marker.

Tumour markers include alpha fetoprotein, apolipoprotein D, BAG-1 (RAP46 protein), CA19-9 (sialyl lewisa), CA50 (carcinoma associated mucin antigen), CA125 (ovarian cancer antigen), CA242 (tumour associated mucin antigen), chromogranin A, clusterin (apolipoprotein J), epithelial membrane antigen, epithelial-related antigen, epithelial specific antigen, gross cystic disease fluid protein-15, hepatocyte specific antigen, heregulin, human gastric mucin, human milk fat globule, MAGE-1, matrix metalloproteinases, melan A, melanoma marker (HMB45), mesothelin, metallothionein, microphthalmia transcription factor (MITF), Muc-1 core glycoprotein, Muc-1 glycoprotein, Muc-2 glycoprotein, Muc-5AC glycoprotein, Muc-6 glycoprotein, myeloperoxidase, Myf-3 (Rhabdomyosarcoma marker), Myf-4 (Rhabdomyosarcoma marker), MyoD1 (Rhabdomyosarcoma marker), myoglobin, nm23 protein, placental alkaline phosphatase, prealbumin, prostate specific antigen, prostatic acid phosphatase, prostatic inhibin peptide, PTEN, renal cell carcinoma marker, small intestinal mucinous antigen, tetranectin, thyroid transcription factor-1, tissue inhibitor of matrix metalloproteinase 1, tissue inhibitor of matrix metalloproteinase 2, tyrosinase, tyrosinase-related protein-1, villin, von Willebrand factor.

Cell cycle associated markers include apoptosis protease activating factor-1, bcl-w , bcl-x, bromodeoxyuridine, CAK (cdk-activating kinase), cellular apoptosis susceptibility protein (CAS), caspase 2, caspase 8, CPP32 (caspase-3), cyclin dependent kinases, cyclin A, cyclin B1, cyclin D1, cyclin D2, cyclin D3, cyclin E, cyclin G, DNA fragmentation factor (N-terminus), Fas (CD95), Fas-associated death domain protein, Fas ligand, Fen-1, IPO-38, Mcl-1, minichromosome maintenance proteins, mismatch repair protein (MSH2), poly (ADP-Ribose) polymerase, proliferating cell nuclear antigen, p16 protein, p27 protein, p34cdc2, p57 protein (Kip2), p105 protein, Stat 1 alpha, topoisomerase I, topoisomerase II alpha, topoisomerase III alpha, topoisomerase II beta.

Neural tissue and tumour markers include alpha B crystallin, alpha-internexin, alpha synuclein, amyloid precursor protein, beta amyloid, calbindin, choline acetyltransferase, excitatory amino acid transporter 1, GAP43, glial fibrillary acidic protein, glutamate receptor 2, myelin basic protein, nerve growth factor receptor (gp75), neuroblastoma marker, neurofilament 68kD, neurofilament 160kD, neurofilament 200kD, neuron specific enolase, nicotinic acetylcholine receptor alpha4, nicotinic acetylcholine receptor beta2, peripherin, protein gene product 9, S-100 protein, serotonin, SNAP-25, synapsin I, synaptophysin, tau, tryptophan hydroxylase, tyrosine hydroxylase, ubiquitin.

Cluster differentiation markers include CD1a, CD1b, CD1c, CD1d, CD1e, CD2, CD3delta, CD3epsilon, CD3gamma, CD4, CD5, CD6, CD7, CD8alpha, CD8beta, CD9, CD10, CD11a, CD11b, CD11c, CDw12, CD13, CD14, CD15, CD15s, CD16a, CD16b, CDw17, CD18, CD19, CD20, CD21, CD22, CD23, CD24, CD25, CD26, CD27, CD28, CD29, CD30, CD31, CD32, CD33, CD34, CD35, CD36, CD37, CD38, CD39, CD40, CD41, CD42a, CD42b, CD42c, CD42d, CD43, CD44, CD44R, CD45, CD46, CD47, CD48, CD49a, CD49b, CD49c, CD49d, CD49e, CD49f, CD50, CD51, CD52, CD53, CD54, CD55, CD56, CD57, CD58, CD59, CDw60, CD61, CD62E, CD62L, CD62P, CD63, CD64, CD65, CD65s, CD66a, CD66b, CD66c, CD66d, CD66e, CD66f, CD68, CD69, CD70, CD71, CD72, CD73, CD74, CDw75, CDw76, CD77, CD79a, CD79b, CD80, CD81, CD82, CD83, CD84, CD85, CD86, CD87, CD88, CD89, CD90, CD91, CDw92, CDw93, CD94, CD95, CD96, CD97, CD98, CD99, CD100, CD101, CD102, CD103, CD104, CD105, CD106, CD107a, CD107b, CDw108, CD109, CD114, CD115, CD116, CD117, CDw119, CD120a, CD120b, CD121a, CDw121b, CD122, CD123, CD124, CDw125,

CD126, CD127, CDw128a, CDw128b, CD130, CDw131, CD132, CD134, CD135, CDw136, CDw137, CD138, CD139, CD140a, CD140b, CD141, CD142, CD143, CD144, CDw145, CD146, CD147, CD148. CDw149, CDw150, CD151, CD152, CD153, CD154, CD155, CD156, CD157, CD158a, CD158b, CD161, CD162, CD163, CD164, CD165, CD166, and TCR-zeta.

5 Other cellular markers include centromere protein-F (CENP-F), giantin, involucrin, lamin A&C [XB10], LAP-70, mucin, nuclear pore complex proteins, p180 lamellar body protein, ran, r, cathepsin D, Ps2 protein, Her2-neu, P53, S100, epithelial marker antigen (EMA), TdT, MB2, MB3, PCNA, and Ki67.

10 Cell containing samples may be stained using dyes or stains, or histochemicals, that directly react with the specific biomarkers or with various types of cells or subcellular compartments. Not all stains are compatible. Therefore the type of stains employed and their sequence of application should be well considered, but can be readily determined by one of skill in the art. Such histochemicals may be chromophores detectable by transmittance microscopy or fluorophores detectable by fluorescence microscopy. In general, a cell containing samples may be incubated with a solution comprising at least one histochemical, which will directly react with or bind to chemical groups of the target. Some histochemicals must be co-incubated with a mordant, or metal, in order to allow staining. A cell containing sample may be incubated with a mixture of at least one histochemical that stains a component of interest and another histochemical that acts as a counterstain and binds a region outside the component of interest.
15
20 Alternatively, mixtures of multiple probes may be used in the staining, and provide a way to identify the positions of specific probes.

The following, non-limiting list provides exemplary chromophores that may be used as histological stains or counterstains and their target cells, subcellular compartments, or cellular components: Eosin (alkaline cellular components, cytoplasm), Hematoxylin (nucleic acids),
25 Orange G (red blood, pancreas, and pituitary cells), Light Green SF (collagen), Romanowsky-Giemsa (overall cell morphology), May-Grunwald (blood cells), Blue Counterstain (Trevigen), Ethyl Green (CAS) (amyloid), Feulgen-Naphthol Yellow S (DNA), Giemsa (differentially stains various cellular compartments), Methyl Green (amyloid), pyronin (nucleic acids), Naphthol-
30 Yellow (red blood cells), Neutral Red (nuclei), Papanicolaou stain (which typically includes a

mixture of Hematoxylin, Eosin Y, Orange G and Bismarck Brown mixture (overall cell morphology), Red Counterstain B (Trevigen), Red Counterstain C (Trevigen), Sirius Red (amyloid), Feulgen reagent (pararosanilin) (DNA), Gallocyanin chrom-alum (DNA), Gallocyanin chrom-alum and Naphthol Yellow S (DNA), Methyl Green-Pyronin Y (DNA), Thionin-Feulgen reagent (DNA), Acridine Orange (DNA), Methylene Blue (RNA and DNA), Toluidine Blue (RNA and DNA), Alcian blue (carbohydrates), Ruthenium Red (carbohydrates), Sudan Black (lipids), Sudan IV (lipids), Oil Red-O (lipids), Van Gieson's trichrome stain (acid fuchsin and picric acid mixture) (muscle cells), Masson trichrome stain (hematoxylin, acid fuchsin, and Light Green mixture) (stains collagen, cytoplasm, nucleoli differently), Aldehyde Fuchsin (elastin fibers), and Weigert stain (differentiates reticular and collagenous fibers). A comprehensive list of such stains, their description, and general use is given in R.D. Lillie, "Conn's Biological Stains", 8th ed., Williams and Wilkins Company, Baltimore, Maryland (1969). Suitable mordants and compositions of the preceding are well-known to one of skill in the art.

The following, non-limiting list provides exemplary fluorescent histological stains and their target cells, subcellular compartments, or cellular components if applicable: 4',6-diamidino-2-phenylindole (DAPI) (nucleic acids), Eosin (alkaline cellular components, cytoplasm), Hoechst 33258 and Hoechst 33342 (two bisbenzimidazoles) (nucleic acids), Propidium Iodide (nucleic acids), Spectrum Orange (nucleic acids), Spectrum Green (nucleic acids), Quinacrine (nucleic acids), Fluorescein-phalloidin (actin fibers), Chromomycin A 3 (nucleic acids), Acriflavine-Feulgen reaction (nucleic acid), Auramine O-Feulgen reaction (nucleic acids), Ethidium Bromide (nucleic acids). Nissl stains (neurons), high affinity DNA fluorophores such as POPO, BOBO, YOYO and TOTO and others, and Green Fluorescent Protein fused to DNA binding protein, such as histones, ACMA, Quinacrine and Acridine Orange.

A wide variety of proprietary fluorescent organelle-specific probes are available from Molecular Probes (Eugene, OR), which include mitochondria-specific probes (MitoFluor and MitoTracker dyes), endoplasmic reticulum (ER) and Golgi probes (ER-Tracker and various ceramide conjugates), and lysosomal probes (LysoTracker dyes). These probes, as well as many nonproprietary fluorescent histochemicals, are available from and extensively described in the

Handbook of Fluorescent Probes and Research Products 8th Ed. (2001), available from Molecular Probes, Eugene, OR.

Each cell containing sample may be co-incubated with appropriate substrates for an enzyme that is a cellular component of interest and appropriate reagents that yield colored precipitates at the sites of enzyme activity. Such enzyme histochemical stains are specific for the particular target enzyme. Staining with enzyme histochemical stains may be used to define a subcellular component or a particular type of cell. Alternatively, enzyme histochemical stains may be used diagnostically to quantitate the amount of enzyme activity in cells. A wide variety of enzymatic substrates and detection assays are known and described in the art, and some selected methods are exemplified below.

Acid phosphatases may be detected through several methods. In the Gomori method for acid phosphatase, a cell preparation is incubated with glycerophosphate and lead nitrate. The enzyme liberates phosphate, which combines with lead to produce lead phosphate, a colorless precipitate. The tissue is then immersed in a solution of ammonium sulfide, which reacts with lead phosphate to form lead sulfide, a black precipitate. Alternatively, cells may be incubated with a solution comprising pararosanilin-HCl, sodium nitrite, naphthol ASB1 phosphate (substrate), and veronal acetate buffer. This method produces a red precipitate in the areas of acid phosphatase activity. Owing to their characteristic content of acid phosphatase, lysosomes can be distinguished from other cytoplasmic granules and organelles through the use of this assay.

Dehydrogenases may be localized by incubating cells with an appropriate substrate for the species of dehydrogenase and tetrazole. The enzyme transfers hydrogen ions from the substrate to tetrazole, reducing tetrazole to formazan, a dark precipitate. For example, NADH dehydrogenase is a component of complex I of the respiratory chain and is localized predominantly to the mitochondria.

Other enzymes for which well-known staining techniques have been developed, and their primary cellular locations or activities, include but are not limited to the following: ATPases

(muscle fibers), succinate dehydrogenases (mitochondria), cytochrome c oxidases (mitochondria), phosphorylases (mitochondria), phosphofructokinases (mitochondria), acetyl cholinesterases (nerve cells), lactases (small intestine), leucine aminopeptidases (liver cells), myodenylate deaminases (muscle cells), NADH diaphorases (erythrocytes), and sucrases (small intestine).

Immunohistochemistry is among the most sensitive and specific histochemical techniques. Each histospot may be combined with a labeled binding composition comprising a specifically binding probe. Various labels may be employed, such as fluorophores, or enzymes which produce a product which absorbs light or fluoresces. A wide variety of labels are known which provide for strong signals in relation to a single binding event. Multiple probes used in the staining may be labeled with more than one distinguishable fluorescent label. These color differences provide a way to identify the positions of specific probes. The method of preparing conjugates of fluorophores and proteins, such as antibodies, is extensively described in the literature and does not require exemplification here.

Although there are at least 120,000 commercially available antibodies, the following lists some exemplary primary antibodies known to specifically bind cellular components and which are presently employed as components in immunohistochemical stains used for research and, in limited cases, for diagnosis of various diseases. Anti-estrogen receptor antibody (breast cancer), anti-progesterone receptor antibody (breast cancer), anti-p53 antibody (multiple cancers), anti-Her-2/neu antibody (multiple cancers), anti-EGFR antibody (epidermal growth factor, multiple cancers), anti-cathepsin D antibody (breast and other cancers), anti-Bcl-2 antibody (apoptotic cells), anti-E-cadherin antibody, anti-CA125 antibody (ovarian and other cancers), anti-CA15-3 antibody (breast cancer), anti-CA19-9 antibody (colon cancer), anti-c-erbB-2 antibody, anti-P-glycoprotein antibody (MDR, multi-drug resistance), anti-CEA antibody (carcinoembryonic antigen), anti-retinoblastoma protein (Rb) antibody, anti-ras oncoprotein (p21) antibody, anti-Lewis X (also called CD15) antibody, anti-Ki-67 antibody (cellular proliferation), anti-PCNA (multiple cancers) antibody, anti-CD3 antibody (T-cells), anti-CD4 antibody (helper T cells), anti-CD5 antibody (T cells), anti-CD7 antibody (thymocytes, immature T cells, NK killer cells), anti-CD8 antibody (suppressor T cells), anti-CD9/p24 antibody (ALL), anti-CD10 (also called

CALLA) antibody (common acute lymphoblastic leukemia), anti-CD11c antibody (Monocytes, granulocytes, AML), anti-CD13 antibody (myelomonocytic cells, AML), anti-CD14 antibody (mature monocytes, granulocytes), anti-CD15 antibody (Hodgkin's disease), anti-CD19 antibody (B cells), anti-CD20 antibody (B cells), anti-CD22 antibody (B cells), anti-CD23 antibody (activated B cells, CLL), anti-CD30 antibody (activated T and B cells, Hodgkin's disease), anti-CD31 antibody (angiogenesis marker), anti-CD33 antibody (myeloid cells, AML), anti-CD34 antibody (endothelial stem cells, stromal tumors), anti-CD35 antibody (dendritic cells), anti-CD38 antibody (plasma cells, activated T, B, and myeloid cells), anti-CD41 antibody (platelets, megakaryocytes), anti-LCA/CD45 antibody (leukocyte common antigen), anti-CD45RO antibody (helper, inducer T cells), anti-CD45RA antibody (B cells), anti-CD39, CD100 antibody, anti-CD95/Fas antibody (apoptosis), anti-CD99 antibody (Ewings Sarcoma marker, MIC2 gene product), anti-CD106 antibody (VCAM-1; activated endothelial cells), anti-ubiquitin antibody (Alzheimer's disease), anti-CD71 (transferrin receptor) antibody, anti-c-myc (oncoprotein and a hapten) antibody, anti-cytokeratins (transferrin receptor) antibody, anti-vimentins (endothelial cells) antibody (B and T cells), anti-HPV proteins (human papillomavirus) antibody, anti-kappa light chains antibody (B cell), anti-lambda light chains antibody (B cell), anti-melanosomes (HMB45) antibody (melanoma), anti-prostate specific antigen (PSA) antibody (prostate cancer), anti-S-100 antibody (melanoma, salivary, glial cells), anti-tau antigen antibody (Alzheimer's disease), anti-fibrin antibody (epithelial cells), anti-keratins antibody, anti-cytokeratin antibody (tumor), anti-alpha-catenin (cell membrane), and anti-Tn-antigen antibody (colon carcinoma, adenocarcinomas, and pancreatic cancer).

Fluorophores that may be conjugated to a primary antibody include but are not limited to Fluorescein, Rhodamine, Texas Red, Cy2, Cy3, Cy5, VECTOR Red, ELF.TM. (Enzyme-Labeled Fluorescence), Cy0, Cy0.5, Cy1, Cy1.5, Cy3, Cy3.5, Cy5, Cy7, FluorX, Calcein, Calcein-AM, CRYPTOFLUOR.TM.'S, Orange (42 kDa), Tangerine (35 kDa), Gold (31 kDa), Red (42 kDa), Crimson (40 kDa), BHMP, BHDMAP, Br-Oregon, Lucifer Yellow, Alexa dye family, N-[6-(7-nitrobenz-2-oxa-1, 3-diazol-4-yl)amino]caproyl] (NBD), BODIPY.TM., boron dipyrromethene difluoride, Oregon Green, MITOTRACKER.TM. Red, DiOC.sub.7 (3), DiIC.sub.18, Phycoerythrin, Phycobiliproteins BPE (240 kDa) RPE (240 kDa) CPC (264 kDa) APC (104 kDa), Spectrum Blue, Spectrum Aqua, Spectrum Green, Spectrum Gold, Spectrum Orange,

Spectrum Red, NADH, NADPH, FAD, Infra-Red (IR) Dyes, Cyclic GDP-Ribose (cGDPR), Calcofluor White, Lissamine, Umbelliferone, Tyrosine and Tryptophan. A wide variety of other fluorescent probes are available from and/or extensively described in the *Handbook of Fluorescent Probes and Research Products* 8th Ed. (2001), available from Molecular Probes, Eugene, OR., as well as many other manufacturers.

Further amplification of the signal can be achieved by using combinations of specific binding members, such as antibodies and anti-antibodies, where the anti-antibodies bind to a conserved region of the target antibody probe, particularly where the antibodies are from different species. Alternatively specific binding ligand-receptor pairs, such as biotin-streptavidin, may be used, where the primary antibody is conjugated to one member of the pair and the other member is labeled with a detectable probe. Thus, one effectively builds a sandwich of binding members, where the first binding member binds to the cellular component and serves to provide for secondary binding, where the secondary binding member may or may not include a label, which may further provide for tertiary binding where the tertiary binding member will provide a label.

The secondary antibody, avidin, strepavidin or biotin are each independently labeled with a detectable moiety, which can be an enzyme directing a colorimetric reaction of a substrate having a substantially non-soluble color reaction product, a fluorescent dye (stain), a luminescent dye or a non-fluorescent dye. Examples concerning each of these options are listed below.

In principle, any enzyme that (i) can be conjugated to or bind indirectly to (e.g., via conjugated avidin, strepavidin, biotin, secondary antibody) a primary antibody, and (ii) uses a soluble substrate to provide an insoluble product (precipitate) could be used.

The enzyme employed can be, for example, alkaline phosphatase, horseradish peroxidase, beta-galactosidase and/or glucose oxidase; and the substrate can respectively be an alkaline phosphatase, horseradish peroxidase, beta.-galactosidase or glucose oxidase substrate.

Alkaline phosphatase (AP) substrates include, but are not limited to, AP-Blue substrate (blue precipitate, Zymed catalog p. 61); AP-Orange substrate (orange, precipitate, Zymed), AP-Red substrate (red, red precipitate, Zymed), 5-bromo, 4-chloro, 3-indolylphosphate (BCIP substrate, turquoise precipitate), 5-bromo, 4-chloro, 3-indolyl phosphate/nitroblue tetrazolium/iodonitrotetrazolium (BCIP/INT substrate, yellow-brown precipitate, Biomeda), 5-bromo, 4-chloro, 3-indolylphosphate/nitroblue tetrazolium (BCIP/NBT substrate, blue/purple), 5-bromo, 4-chloro, 3-indolyl phosphate/nitroblue tetrazolium/iodonitrotetrazolium (BCIP/NBT/INT, brown precipitate, DAKO, Fast Red (Red), Magenta-phos (magenta), Naphthol AS-BI-phosphate (NABP)/Fast Red TR (Red), Naphthol AS-BI-phosphate (NABP)/New Fuchsin (Red), Naphthol AS-MX-phosphate (NAMP)/New Fuchsin (Red), New Fuchsin AP substrate (red), p-Nitrophenyl phosphate (PNPP, Yellow, water soluble), VECTORTM Black (black), VECTORTM. Blue (blue), VECTORTM. Red (red), Vega Red (raspberry red color).

Horseradish Peroxidase (HRP, sometimes abbreviated PO) substrates include, but are not limited to, 2,2' Azino-di-3-ethylbenz-thiazoline sulfonate (ABTS, green, water soluble), aminoethyl carbazole, 3-amino, 9-ethylcarbazole AEC (3A9EC, red). Alpha-naphthol pyronin (red), 4-chloro-1-naphthol (4C1N, blue, blue-black), 3,3'-diaminobenzidine tetrahydrochloride (DAB, brown), ortho-dianisidine (green), o-phenylene diamine (OPD, brown, water soluble), TACS Blue (blue), TACS Red (red), 3,3',5,5'-Tetramethylbenzidine (TMB, green or green/blue), TRUE BLUETM. (blue), VECTORTM. VIP (purple), VECTORTM. SG (smoky blue-gray), and Zymed Blue HRP substrate (vivid blue).

Glucose oxidase (GO) substrates, include, but are not limited to, nitroblue tetrazolium (NBT, purple precipitate), tetranitroblue tetrazolium (TNBT, black precipitate), 2-(4-iodophenyl)-5-(4-nitrophenyl)-3-phenyltetrazolium chloride (INT, red or orange precipitate), Tetrazolium blue (blue), Nitrotetrazolium violet (violet), and 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT, purple). All tetrazolium substrates require glucose as a co-substrate. The glucose gets oxidized and the tetrazolium salt gets reduced and forms an insoluble formazan which forms the color precipitate.

Beta-galactosidase substrates, include, but are not limited to, 5-bromo-4-chloro-3-indoyl beta-D-galactopyranoside (X-gal, blue precipitate). The precipitates associated with each of the substrates listed have unique detectable spectral signatures (components).

5 The enzyme can also be directed at catalyzing a luminescence reaction of a substrate, such as, but not limited to, luciferase and aequorin, having a substantially non-soluble reaction product capable of luminescing or of directing a second reaction of a second substrate, such as but not limited to, luciferine and ATP or coelenterazine and Ca.sup.++, having a luminescing product.

10 The following references, which are incorporated herein provide additional examples. J.M Elias (1990) *Immunohistopathology: A practical approach to diagnosis*. ASCP Press (American Society of Clinical Pathologists), Chicago; J. F. McGinty, F. E. Bloom (1983) *Double immunostaining reveals distinctions among opioid peptidergic neurons in the medial basal hypothalamus*. Brain Res. 278: 145-153; and T. Jowett (1997) *Tissue In situ Hybridization: Methods in Animal Development*. John Wiley & Sons, Inc., New York; J Histochem Cytochem 1997 December 45(12):1629-1641.

15 Cellular preparations may be subjected to *in-situ* hybridization (ISH). In general, a nucleic acid sequence probe is synthesized and labeled with either a fluorescent probe or one member of a ligand:receptor pair, such as biotin/avidin, labeled with a detectable moiety. Exemplary probes and moieties are described in the preceding section. The sequence probe is complementary to a target nucleotide sequence in the cell. Each cell or cellular compartment containing the target nucleotide sequence may bind the labeled probe. Probes used in the analysis may be either DNA or RNA oligonucleotides or polynucleotides and may contain not only naturally occurring nucleotides but their analogs such as dioxynin dCTP, biotin dCTP 7-azaguanosine, azidothymidine, inosine, or uridine. Other useful probes include peptide probes and analogues thereof, branched gene DNA, peptidomimetics, peptide nucleic acids, and/or antibodies. Probes should have sufficient complementarity to the target nucleic acid sequence of interest so that stable and specific binding occurs between the target nucleic acid sequence and the probe. The degree of homology required for stable hybridization varies with the stringency

20

25

30

of the hybridization. Conventional methodologies for ISH, hybridization and probe selection are described in Leitch, et al. *In Situ Hybridization: a practical guide*, Oxford BIOS Scientific Publishers, Microscopy Handbooks v. 27 (1994); and Sambrook, J., Fritsch, E.F., Maniatis, T., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989).

5

The present invention is further illustrated by the following examples, which should not be construed as limiting in any way. The contents of all cited references are hereby expressly incorporated by reference.

Example 1. Construction of Tissue Microarrays for a Survival Analysis of The Estrogen Receptor (ER) and HER2/neu and for Analysis of Nuclear Associated Beta-catenin

Tissue microarray design: Paraffin-embedded formalin-fixed specimens from 345 cases of node-positive invasive breast carcinoma were identified. Areas of invasive carcinoma, away from *in situ* lesions and normal epithelium, were identified and three 0.6cm punch "biopsy" cores were taken from separate areas. Each core was arrayed into a separate recipient block, and five-micron thick sections were cut and processed as previously described (Konenen, J. et al., Tissue microarrays for high-throughput molecular profiling of tumor specimens, (1987) *Nat. Med.* 4:844-7). Similarly, 310 cases of colon carcinoma were obtained and arrayed, as previously described (Chung, G.G. et al., *Clin. Cancer Res.* (In Press)).

20

Immunohistochemistry: Pre-cut paraffin-coated tissue microarray slides were deparaffinized and antigen-retrieved by pressure-cooking (Kato, A.K. et al., (1997) *Biotech Histochem.* F2:291-8). Slides were stained with antibodies to one of three target antigens: monoclonal anti-E.R. (mouse, Dako Corporation, Carpinteria, CA), polyclonal anti-HER2/neu (rabbit, Dako Corp.), monoclonal (mouse clone 14, BD Transduction Labs, San Diego CA) anti-beta-catenin, or polyclonal rabbit anti-beta-catenin. Primaries were incubated overnight at 4°C. A corresponding goat antimouse or anti-rabbit secondary antibody conjugated to a horseradish peroxidase decorated dextran-polymer backbone was then applied for 1 hr (Envision, DAKO Corp.). Target antigens were either visualized with a visible light chromagen (Diaminobenzidine, DAKO) for visual analysis, or a fluorescent chromagen (Cy-5-tyramide, NEN Life Science Products, Boston, MA). Slides designated for automated analysis were

25

30

counterstained with DAPI for visualization of nuclei, and either polyclonal rabbit anticytokeratin (Zymed, So. San Francisco, CA) or rabbit anti-alpha-catenin (?) to distinguish between tumor cells and stroma as well as to visualize the cell membrane. In many cases, exponentially subtracted images of histospots stained with anti-cytokeratin provided an acceptable marker for the cell membrane due to the sub-membranous coalescence of cytokeratin in tumor cells. These antibodies were visualized using either Cy3- or Alexa 488-conjugated goat anti-mouse or anti-rabbit secondary antibodies (Amersham, Piscataway, NJ and Molecular Probes, Eugene, OR). Slides designated for visual inspection were counterstained with ammonium hydroxide acidified hematoxylin. Manual examination of microarrays for E.R., HER2/neu, and beta-catenin levels has been previously described (Snead, D.R. et al.,(1993) *Histopathology* 23:233-8).

Image analysis: Images of microarrays were obtained using a Deltavision platform and software (SoftWorx 2.5) (Applied Precision, Issaquah, WA), with an attached water-cooled Photometrics series 300 camera through a 10x Nikon Super-Fluor lens on a TE200 inverted fluorescent microscope with automated X,Y,Z stage movement. Low power images of microarrays were stitched together using multiple (~1500) low resolution images of the microarray (64x64 pixel). These images were analyzed by software algorithms described herein to determine the location of each . Subsequently, monochromatic, high resolution (1024x1024 pixel) images were obtained of each , both in the plane of focus and 8 microns below it. Image pairs for each fluorescent dye were obtained. Images were analyzed using additional algorithms as follows, in brief. Regions of interest (tumor) were identified using a mask derived from a ubiquitously-expressed epithelial-specific antigen (either cytokeratin or alpha-catenin). Images of fluorescently-tagged membrane and nuclear compartments were exponentially subtracted until a set amount of image intensity remained. Images were then combined so that there was minimal overlap of signal from one compartment to the next. Pixels in which a significant degree of overlap was present were negated from further analysis. The pixel intensity of exponentially subtracted images of the target antigen were assigned to one of three compartments: nuclear, membrane, or non-nuclear non-membrane (cytoplasm). Target intensities were analyzed as described below. For E.R. only nuclear-localized signal was used, for HER2/neu only membrane-localized signal was analyzed. For beta-catenin total signal, the ratio of nuclear to membrane signal, and the ratio of nuclear to total signal was analyzed.

Data analysis: staining scores from the breast cancers represent the averaged (for ER) or maximized (for HER2/neu) results from two scorable histospots. Subsequent studies revealed that analysis of a single histospot could provide significant statistical power to judge outcomes, so that staining scores from the colon cancer array represent the result of only one histospot.

Overall survival analysis was assessed using Kaplan-Meier analysis and the Mantel-Cox log rank score for assessing statistical significance. Relative risk was assessed using the univariate Coxproportional hazards model. Analyses were performed using Statview 5.0.1 (SAS Institute, Cary NC).

10

11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210